Comparative Evaluation of a New Lactation Curve Model for Pasture-based Holstein-Friesian Dairy Cows

S. A. Adediran,[†¶] D. A. Ratkowsky,[†] D. J. Donaghy,[‡β] and A. E. O. Malau-Aduli[†*]

[†]Animal Production and Genetics, School of Agricultural Science/Tasmanian Institute of Agriculture, University of Tasmania, Private Bag 54 Hobart, TAS 7001, Australia.

[‡]Dairy Research Centre, Tasmanian Institute of Agriculture, P. O. Box 3532 Burnie, TAS 7320.

[*]Corresponding author: A. E. O. Malau-Aduli, Tel: +61-3-6226-2717, Fax: +61-3-6226-2642

E-mail: **Aduli.MalauAduli@utas.edu.au**, **aduli40@yahoo.co.uk**

Present address: [β]Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Private Bag 11-222, Palmerston North 4442, New Zealand.

[¶]Genetic Epidemiology, Dalla Lana School of Public Health, University of Toronto 155 College Street, Toronto, ON M5T 3M7, Canada.

INTERPRETATIVE SUMMARY

COMPARATIVE EVALUATION OF LACTATION MODELS – Adediran et al.

Fourteen lactation models were compared with two forms of a new log-quadratic model utilizing test-day milk yield data from pasture-based Holstein-Friesian cows. The goodness of fit of all but two of the models was similar for average lactation, but differed in predicting initial, peak and total milk yields. The new model was more parsimonious and performed better than some of the existing models. It is therefore proposed for modelling lactation in pasture-based dairy cows.

ABSTRACT

Fourteen lactation models were fitted to average and individual cow's lactation data from pasture-based dairy systems in the Australian states of Victoria and Tasmania. The models included a new "log-quadratic" model and a major objective was to evaluate and compare the performance of this model with the other models. Nine empirical and five mechanistic models were first fitted to average test-day milk yield of Holstein-Friesian dairy cows using the non-linear procedure in SAS. Two additional semi-parametric models were fitted using a linear model in AsReml. To investigate the influence of days to first test day and the number of test days, five of the best-fitting models were then fitted to individual cow lactation data. Model goodness of fit was evaluated using criteria such as the residual mean square, the distribution of residuals, the correlation between actual and predicted values and the Wald-Wolfowitz runs test. Goodness of fit was similar in all but one of the models in terms of fitting average lactation but they differed in their ability to predict individual lactations. In particular, the widely-used incomplete gamma model was the model that most displayed this failing. The new log-quadratic model was robust in

fitting average and individual lactations, less affected by sampled data and more parsimonious in having only three parameters, each of which lends itself to biological interpretation.

Key Words: Lactation model, pasture-based, dairy cows, test-day milk yield.

INTRODUCTION

In Australia, dairy cows rely on fodder from pasture for about 70% of feed energy (Dairy Australia 2008). Compared to stall-fed cows, milk yield patterns in pasture-based dairy cows are more prone to fluctuations due to seasonality of pasture production, pasture management practices, undetected sub-clinical diseases, nutritional interventions and other management practices used to mitigate feed shortfalls (Kolver and Muller 1998, Olori et al. 1999, Collard *et al.* 2000, Tekerli *et al.* 2000). Short lactations arising from synchronised calving patterns to match pasture availability, occasional occurrence of double peaks and irregular milk recording logistics are also some peculiar features of pasture-based dairy systems. Under such varying environmental conditions, high genetic merit cows are more likely to exhibit depressed milk yields (Kolver and Muller 1998, Olori et al. 1999).

The lactation curve which can be modelled using mathematical functions (Beever et al. 1991, Schaeffer 2004), generally takes the shape of an increase to a peak 4-8 weeks into lactation, followed by a gradual decline until drying up. These functions have the advantage of minimizing random variation while simultaneously summarising the lactation profile into biologically interpretable parameters. The resulting curve parameter estimates can be further analyzed to predict future yields from incomplete lactation records, detect deviation of an individual cow or a herd of cows from the expected performance, provide early estimates of 305-day milk yields for breeding decisions (Jensen 2001, Schaeffer 2004) and to monitor responses to changes in management factors (Morant and Gnanasakthy 1989, Pollott 2000). However, variation among individual cows (Olori et al. 1999), data properties (Macciotta et al. 2005) and the aforementioned peculiarities accentuate differences in curve patterns hitherto referred to as irregular lactations (Macciotta et al. 2005). It is desirable to have a model, which is easy to apply, biologically meaningful, less constrained by atypical lactations and suitable for describing short lactations (Pollott and Gootwine 2000) to reduce milk-recording costs. For the first time, the new log quadratic (LQ) model is being proposed because of its peculiar ability to fit both inclining and declining lactation rates from initial milk yield, thus being less constrained by *a priori* assumption of an incline to peak yield that all other models make.

The mathematical functions available to model lactation profiles are many, and include empirical (linear or non-linear), mechanistic, and semi-parametric types (Schaeffer 2004, Sherchand et al**.** 1995). The incomplete gamma (**IG**) function (Wood 1967), is the most widely used to model the entire lactation in dairy cows. Empirical models, often criticized for overestimation of milk yield in early lactation, for having pre-determined curve shape and failure to relate curve parameters to mammary gland physiology (Pollott 2000, Macciotta et al. 2005) are still the models of choice by dairy researchers and economists (Tozer and Huffaker 1999) for their relative ease of application and good fitness to diverse lactations.

Although mechanistic models (Neal and Thornley 1983, Pollott 2000, Grossman and Koops 2003) offer insight into the mammary gland physiological processes such as parenchyma cell proliferation, differentiation and apoptosis (Knight and Wilde 1993, Knight et al. 1998), they are often over-parameterized and fit data poorly based on current monthly milk recording systems (Pollott 2000). For their flexibility in fitting time-series for events with various curves, semi-parametric functions such as Legendre polynomials (Kirkpatrick et al. 1994) and cubic splines (White et al. 1999) are also suitable for lactation modeling.

Variation in the goodness of fit of lactation models in diverse production systems has been reported (Tozer and Huffaker 1999, Macciotta et al 2005). Olori *et al.* (1999) observed that the polynomial function (Ali and Schaeffer 1987) was the best of five models compared in a farm-based study while Garcia and Holmes (2001) found no difference in the fit of di-phasic and linear-based split-plot models for pasture-based Holstein Friesian cows. Papajcsik and Bodero (1988) evaluated twenty lactation models and concluded that the IG model ($Y_t = at^b e^{-ct}$) and its derivative $Y_t = at^b / \cosh(ct)$ best fitted the data from cows in a sub-tropical environment. In comparison, Val-Arreola *et al.* (2004) fitted five lactation models to data from small-scale and intensive dairy systems in Mexico and reported that the mechanistic model proposed by Dijkstra *et al.* (1997) gave statistically significant parameter estimates and the lowest residual mean squares, while Silvestre et al. (2006) evaluated seven functions, including three Legendre polynomial and cubic splines in stall-feeding systems and concluded that the spline model best fitted the lactation data. In addition to variation in individual animal production pattern (Olori et al. 1999) and differences in production systems (Tozer and Huffaker 1999, Val-Arreola et al. 2004), the mathematical property (Macciotta et al. 2005) as well as lactation data properties such as day at first test-day, number of available records and the interval between test-days (Berry et al. 2005, Silvestre et al. 2006) have been shown to affect the goodness of fit of a model

The objectives of this paper were to evaluate the goodness of fit of a new log-quadratic (**LQ**) model for fitting average and individual cow's lactation and to compare it with the current state-of-the-art lactation models. Further we evaluated the effect of days to first test-day and number of test-day records on the goodness of fit of the five best-fitting models.

MATERIALS AND METHODS

Sites, Climatic Conditions and Production Systems

The two data sets used in the study were from the Australian states of Tasmania and Victoria which are similar in climatic conditions. In addition, the production system in southeast Gippsland, Victoria, is similar to that used in dairying in Tasmania. Average maximum temperatures are 30°C (86°F) and 21°C (70° F) in Victoria and Tasmania, respectively, in summer (December – February) and 15°C (59°F) and 12°C (54°F), respectively, in winter (June – August). The annual rainfall varies from 626 mm to 2,400 mm. Most dairy herds in Tasmania are located in the North-West while the research farm supplying DATA2 is located in the southeast of Victoria. The relatively mild winter in both sites offer unique opportunities for year-round grazing where precipitation is not limiting. Dairying in both states is characterized by seasonal, low-input pasture-based milk production. Productivity increases are achieved through an increasing use of hay and grain supplements. The Holstein-Friesian (**FF**) breed constitutes about 70 percent of the dairy cows milked in both states.

Daily milk yields are automatically recorded into a database through on-line milking machines for each individually tagged cow. Test-day milk yield records are morning milk yields, collected once a month on average for each registered herd. Evening milk yield are estimated based on standards determined by the Australian Dairy Herd Improvement Scheme (ADHIS 1999). Milk yields from sick or mastitic cows are not collected and treated as missing for the month.

Data Management

Two data sets (DATA1 and DATA2) of FF cows from the Australian states of Tasmania and Victoria respectively were used in the study. DATA1 comprised 76,760 records (9,505 lactations) from 154 dairy herds collected from 2005-2007, while DATA2 from a single dairy research farm consisted of 19,987 records (2,138 lactations) from 1998-2005. The data were edited to exclude cows with; incorrect or missing birth or calving dates, lactation length <100 or >305 days, and less than five test-day records for a particular lactation. Records of cows with first post-partum recorded day in milk (**DIM**) less than 4 days or greater than 120 days and greater than parity 5 were also excluded from the analysis. Parities >2 were pooled and referred to as parity 3. Lactation stage in months (test-day) was obtained as the number of days from calving following the first fifteen days post-partum which was taken as the first test-day post-partum. Seven age groups were defined in DATA1 as follows; $18 \leq \text{age} < 24$, $25 \leq \text{age} < 30$, $31 \leq \text{age} < 36$, $37 \leq \text{age} < 42$, $43 \leq \text{age} < 54$, $55 \leq \text{age} < 67$, $68 \leq \text{age} < 75$ in age classes 1 - 7 and number of records 11, 947, 6, 890, 8, 426, 7, 540, 14, 890, 12, 867, 10, 241 respectively. Summary statistics for herd characteristics are presented in Table 1.

Lactation Models

The various lactation models used to evaluate test-day milk yield (L/d) of the FF cow are shown in Table 2. We propose a new second-degree polynomial model, subsequently referred to as the "log quadratic" (LQ) model, for modeling lactation in dairy cows. Expressed in its general form, the second degree polynomial is a parabola with the equation $LogY_t = a_1x^2 + a_2x + a_3$, where x is log-transformed time such as DIM, but it can be written in a standard or vertex form as

$$LogY_t = a\,(b - Logt)^2 + c \qquad 1$$

where $LogY_t$ is the $\log_e$ transformed test-day milk yield, $Logt$ is the $\log_e$-transformed time t in days, weeks or months in milk and $a \neq 0$, b and c are parameters of the model. Parameter a controls the rate of incline to the peak and the rate of post-peak decline, b is the log-transformed day at peak milk yield and c is the log-transformed peak milk yield. Parameter b not only is the value of Log$t$ at which maximum milk production occurs, it is also the axis of symmetry of the parabola.

To facilitate equivalent comparison of the LQ with and the IG and modified gamma (MG, Morant and Gnanasakthy 1989) models the LQ was also fitted to test-day milk yield directly (i.e., in an untransformed form) as

$$Y_t = exp\ (\mathrm{a}\ (\mathrm{b} - \mathrm{Log}t)^2 + \mathrm{c})$$

Insert Table 1 here

Statistical Analysis

Average Lactation

In order to determine the average lactation curve of pasture-based FF cows, the test-day milk yield of DATA1 was adjusted using the mixed models procedure (PROC MIXED) of SAS (SAS 2002), according to the model;

$$Y_{ijklmno} = \mathrm{L_i} + H_j + TD_k + CY_l + P_m + \mathrm{AGE_n} + e_{ijklmno} \qquad 3$$

where $Y_{ijklmno}$ is the $_{ijklmno}{}^{th}$ observation on test-day milk yield of lactation i with fixed effects; H_j of jth herd (j = 1, 2...154), TD_k of kth test-day (k = 1, 2…10) being 15, 45… 285 days in milk, CY_l of lth calving year (l = 1, 2, 3), P_m of mth parity (m = 1, 2, 3) and AGE_n (n = 1…7). The lactation effect (L) was treated as a random effect nested within herd and $e_{ijklmno}$ is a random sampling effect of the lactation with mean zero and variance σ^2_e. Multiple lactations on the same cow were assumed uncorrelated. Interaction terms were initially included in the model but proved to be not significant and were subsequently dropped.

Least squares means of test-day milk yield were obtained from this model and used in fitting the average lactation (AL) curve of the FF cow. Test-day milk yield and test-days (lactation stage) were fitted to each of the lactation models in turn using the Marquardt's iterative method of the non-linear (NLIN) procedure of SAS (SAS 2002). The cubic Spline and Legendre Polynomial models were fitted as linear models in AsReml (Gilmour et al. 2002). To facilitate comparison with the LQ, the IG *(*Wood 1967) was also fitted in log-linear form, while the LQ and the modified *gamma* (**MG**) (Morant and Gnanasakthy et al. 1989) were fitted to test-day milk yield. The fitted empirical models were the; incomplete *gamma* (IG), modified *gamma* (MG), exponential (**EXP**, Wilmink 1987), polynomial (**PL**, Ali and Schaeffer 1987), quadratic polynomial (QP, Dave 1971), parabolic exponential (**PE**, Sikka 1950) and the log-quadratic (**LQ**). The fitted modified empirical functions were the incomplete *gamma* ($\mathbf{IG_L}$), modified *gamma* ($\mathbf{MG_n}$), and the log-quadratic ($\mathbf{LQ_n}$)**.** The fitted mechanistic models were the bi-compartmental (**BC**, Fergusson and Boston 1993), the Dijkstra (**DJ**, Dijkstra et al. 1997), the Pollott (PT) and the modified Pollott ($\mathbf{PT_2}$), while the semi-parametric models included the Legendre Polynomial (**LG**, Kirkpatrick et al. 1994) and the cubic Spline (**SPL**, Green and Silverman 1994). The mathematical functions, source and number of parameters of the models are shown in (Table 2).

The EXP model (Wilmink 1987) was fitted as a three-parameter model with the constant parameter (k) set at 0.46, this being the best fitting value for average mean yield in a preliminary analysis of the data sets, during which the starting values of the non-linear (**NLIN**) procedures were also determined. The parameters of the PT and DJ models were constrained using the bound statement in SAS (bound > 0), otherwise the models failed to converge.

Insert Table 2 here

In all the models Y_t is test-day milk yield in litres per day, at time t (DIM), a, b, c, d, e, α_i and ϕ are parameters that define the scale and shape of the curve, t' (MG and MG_n) = (DIM – 150) / 100, t_1 and t_2 (PL) are t / 305 and 305 / t respectively, n (PT, and PT_2) = t - 150 and k is a constant. In all PT models, parameter a is the maximum milk secretion potential, b and d are proportions of milk yield potential and milk yield loss at parturition respectively, while c and e are the growth and death rate parameters of the two logistic curves respectively. On the other hand, parameters b and d (BC) and b and c (DJ) represent the rate of cell proliferation and death, respectively.

In the Legendre polynomial model (LP), $w = 2\left(\frac{t - t_{\min}}{t_{\max} - t_{\min}}\right) - 1$

where $t_{\min}$ = 15 and $t_{\max}$ = 285.

Other lactation parameters of the AL such as the peak milk yield (**PY**), day at peak yield, total milk yield to 305 day (**TMY**) and lactation persistency were estimated from the curve parameters for each model (Table 4). In order to obtain a uniform and comparable value of persistency across models, persistency was defined here as the ratio of the difference in daily milk yield millilitre per day at DIM 60 and 270 to the number of days during that same period using the formula

$$P_{lact} = (MY_{60} - MY_{270}) / 210 \quad (4)$$

where P_{lact} is the persistency of lactation, MY_{270} and MY_{60} are test-day milk yield on 270 and 60 DIM respectively. Cows with lower P_{lact} values are more persistent than those with higher values. These days were chosen because for most pasture-based dairy cows peak milk yield occurs before or at 60th day post-partum, while lactations would last 270 days or more in typical annual calving systems.

Model Evaluation for Average Lactation

The goodness of fit of each model fitted to the AL was evaluated based on the analysis of residuals. Measures of prediction error including the residual mean square (**RMS**), the magnitude and distribution of residuals represented as the plot of residuals against lactation stage (Figure 2) and the correlation between observed and predicted test-day milk yield were used. The Bayesian Information Criteria (BIC), (Leonard and Hsu, 2001) was used to compare models. Further, the AL parameter estimates of each model (Table 4) were used to predict test-day daily milk yield on the successive 10th day in lactation i.e. DIM = 10, 20, 30 etc. Thus 30 predicted values were obtained for each model. These were used to compute a new residual mean square (**RMS$_W$**) as a measure of goodness of fit for each model using the formula according to Pollott and Gootwine (2000):

$$RMS_W = (\sum_{t=1}^{N} (M_{rPD} - M_{rHC})^2) / (n - N) \quad (5)$$

where M_{rPD} and M_{rHC} were the predicted and AL lactation yields on each DIM, respectively, *n* was the number of test-day milk records (30) in the lactation and N was the number of parameters in the model (see Table 3). The resulting RMS_W was used to rank the models in order of best to worst goodness of fit. The log-transformed models (LQ, MG and IG_L) were compared separately from the other models. This ranking, and the number of parameters in the models were used in selecting five models which were used in the further analysis of individual cows' lactations.

Individual Cow's Lactation

The entire test-day milk yields of individual cows from DATA1 and DATA2 were fitted to the five best-fitting models using the NLIN procedures in SAS as described in the previous section on model evaluation for average lactation. The objective was to determine how each model performs with diverse lactation data. In order to determine the effect of data availability on the goodness of fit of the five models both data sets were partitioned according to the number of post-partum days before the first test-day and the number of test-days.

There were two groups, determined by whether the first test day was less than or greater than 60 days, respectively. Within each of these groups, the data were further sub-divided into three classes, based upon whether there were 5, 6 - 7 or > 7 available lactation records. Peak milk yield in our data occurred on or before day 60 post partum. Thus, DATA1 had six partitions, viz. L60A (first test day <60, number of records = 5), G60A (first test day >60, number of records = 5), L60B (first test-day < 60 and number of records = 6 or 7), and so on down to the sixth partition G60C (first test-day > 60 and number of records > 7). All available data were used in all but the L60B and L60C groups (DATA1) in which a random sample of 500 lactations was taken from 1957 and 6390 lactations, respectively. Due to limitation in data size, DATA2 was partitioned into three groups namely L60C, G60B and G60C. The interval between recorded test-days was not considered, as neither of the data sets had records for all test days (1- 305) or had uniform test-days intervals. The number of lactations per group, and the mean and standard deviation of milk yield, are shown in Table 2.

Model Evaluation for Individual Cow's Lactation

The following criteria were used to compare the goodness of fit of the models fitted to individual lactations and the partitioned data (Tables 6 and 7).

a) Average and standard deviation of error (residuals), which measures the error in absolute terms (Congleton and Everett, 1980) without recognizing its variation through the lactation.

b) Association between actual and predicted milk yield, measured as the proportion of explained variation in the response variables, was calculated as

$$R^2 = 1\text{-}SSE/CSS \qquad 6$$

where SSE was the error sum of squares and CSS was the corrected sum of squares of milk yield.

c) The randomness of the distribution of the errors in serially correlated data can be quantified, measured and tested by the Wald-Wolfowitz (W, non-random distribution for $P < 0.05$) runs tests (Constantinides, 1988). The Wald-Wolfowitz runs test is a nonparametric test applied to the errors of each lactation. A significant test ($P < 0.05$) indicates the presence of longer than expected sequences of positive or negative residuals.

d) Differences between extremes of observed and predicted milk yield represented by the percentage of estimated milk yields ≤ 0 (EXLO) or greater than the highest observed yield (EXHI), (Silvestre et al. 2006) were also used as a measure of goodness of fit. Mean test-day milk yield were 12.7 L/d and 18.9 L/d with only 25 and 32 records having milk yield > 40 L/d in DATA1 and DATA2, respectively. This corresponds to an expectation of 0.03% (DATA1) and 0.16% (DATA2). Predicted milk yield values were examined for yield ≤ 0, which is not expected biologically, or > 40, an expectation higher than observed. In either case the model is considered less reliable. Criteria a), b), and d) were calculated across all records, whereas criteria c) was calculated within lactations. (Tables 6 and 7).

RESULTS

Average Lactation

Model parameter estimates, residual mean square (**RMS**) and Bayesian Information Criteria (**BIC**) of all the fitted models are shown in Table 4. The goodness of fit as determined by the RMS did not differ significantly among the three models fitted with log-transformed milk yield i.e. the modified gamma (MG), the incomplete gamma (IG_L) and the log-quadratic (LQ). However, RMS was significantly higher ($p<0.05$) in the parabolic exponential (PE), and quadratic polynomial (QP) compared with the other empirical models. Similarly, the RMS did not differ among the mechanistic models except the modified Pollott (PT_2), which had lower value. Among the semi-parametric models, the Legendre polynomial (LG) fitted the AL with less error bias than the cubic spline (SPL). The BIC was lower in the log-transformed empirical and the mechanistic models compared with the other models except the EXP and PL. Figure 1 shows the average lactation (AL) profile of the Holstein-Friesian cow with initial, peak and nadir milk yield (L/d) at 12.6, 13.2 and 8.9 respectively. Only the PT_2 and the QP models fitted a continuously declining lactation curve (not shown in figure) in contrast to a curve rising to a peak before the decline. The PL and LQ models most accurately predicted the AL as shown in the plot of residuals (Figure 2).

Insert Figure 1 here

Insert table 3 here

Correlation between predicted and observed milk yields (not shown) was highest in the PL model at 0.997, lowest in the PT model (0.764) and averaged 0.989 for all the models. All the models except LQ had highly correlated parameter estimates (not shown). The residuals derived from fitting the various functions to the AL, shown in Figure 2 were generally random except around peak milk yield. All the models achieved similar accuracies in predicting the AL. (Figure 2).

Table 5 shows the predicted AL initial, peak and nadir milk yield (L/d), and 305d milk yield, lactation persistency and days to peak milk yield values. The MG and LQ models gave initial milk yield values closest to the AL while the IG_L over-predicted initial milk yield by 2.4 litres compared with 0.65 to 0.67 in the MG_n and the LQ_n models respectively. Of the other models only the PL unpredicted initial milk yield. All the models under-predicted peak milk yield by between 0.6 to 1.4 litres per day. The EXP and LEG models most accurately predicted the day on which peak yield occurred. Except the EXP and PT_2 all the modes gave accurate prediction of TMY although predictions were best in both forms of the LQ and MG models (Table 4). All the models predicted lactation persistency within 1.5 to 9.6 mL per day but the LQ, MG, EXP and PL models gave the most accurate prediction.

Individual Cow's Lactation

All the tested models were ranked in order of best to worst goodness of fit based on the comparison of RMS_w values obtained from equation 5. The order was LQ, MG and IG_L for the log-transformed models and MG_n, PL, LQ_n, IG, PE, SPL, QP, BC, DJ, LEG, PT, EXP and PT_2, respectively for the other models. Based on this ranking, number of parameters in the model and the magnitude and distribution of the residuals (Table 4 and Figure 2), three models and the two forms of the LQ model were selected for further tests with individual cow's lactation and for the evaluation of day at first test-day and number of observations on the goodness of fit. Additional consideration for model selection was to include at least one log-transformed, one non-linear and one mechanistic model among the models to be compared with the LQ. The selected models were the LQ and MG, fitted to log-transformed test-day milk yield and the LQ_n, IG, and BC models fitted directly to test-day milk yield. The five models were first fitted to all available data (Table 6) and then to data partitioned on the basis of restricted data.

Comparison of Models fitted to Individual Cow's Lactation

Table 6 shows the results for the error criteria used in assessing the goodness of fit of the five models fitted to the individual cow's lactation (DATA1 and DATA2). Mean error, RMS and their standard errors showed similar trends in the goodness of fit of patterns of all models in both data sets, although as expected the margin of errors was higher in the more variable DATA1. Higher variation in DATA1 was also reflected in the correlation between observed and predicted values. Mean errors were lowest in the LQn and MG models compared to the IG in DATA1. In contrast, RMS was lower for the IG compared to the BC and LQn in DATA2, suggesting that the IG may be more suited to fitting data from a more uniform production pattern.

All the models except MG (DATA1) showed a non-random distribution of errors (W) suggesting longer than expected runs of negative or positive residuals. The proportion of zero or negative test-day milk yields (EXLO) was highest (3.1%) for the IG while the observed milk yields were generally lower than the expectation > 40 (EXHI). The correlation between observed and predicted milk yield as measured by R^2 among models was lower in DATA1 compared to DATA2 and lowest for the IG compared with the other models, except in DATA2 where BC had the lowest R^2 values. Mean error, RMS, EXLO and R^2 were similar for the MG and LQ_n models in both data sets. The randomness of residuals (W) was also similar for both models in DATA2 although the MG had a slightly lower value. The percentage estimated milk yield higher than expectation (EXHI) was higher in DATA2 compared to DATA1 for all models.

Effect of Day at First Test-day and available Records on Model's goodness of fit.

The results of the goodness of fit criteria of the five models fitted to the sampled data are shown in Table 7. The IG model was the most affected by the sampled data irrespective of the group or sub-group. Using IG, lactations were better predicted in the L60 compared with the G60 group, mean error = 0.26 – 0.43 *vs.* 0.47 – 0.57 and RMS = 13.6 – 14.9 *vs.* 17.1 – 20.7, respectively. Similarly, within the L60 sub-groups, mean error was highest in L60A and declined with more available data, whereas mean error increased irrespective of number of observations in the G60 group. Mean error tended to remain stable at 0.005 ± 0.03 and 0.002 ± 0.04 for BC and LQ_n, respectively. Prediction error as determined by RMS was highest for IG and lowest for LQ_n although large standard deviations, 24 – 55 (IG), 10 – 27 (BC) and 8 – 16 (LQ_n), suggest prediction bias in all the models. Correlation between observed and predicted milk yield (R^2) for IG also declined with fewer data in both L60 and G60 groups, declined for BC in the L60D and G60C sub-groups and remained stable for LQ_n. However, the number of lactations with a random distribution (W) had similar effects in all models. Significant p –values of the Wald-Wolfowitz test was highest for the L60A and the G60A sub-groups and smallest for the L60C and G60C sampled sub-groups. Non-randomness of error was similar in all data partitioned groups although the pattern was lower in the G60D group. The mean error of the BC and LQn was lower than that for the IG partition group. Both models were equally affected by smaller data size irrespective of day and first test-day. LQ tended to have the least variation (lower SD) than the other models. The number of available test-day records also affected the R^2 value more severely for IG than for the BC and LQn which tended to maintain stable correlation between observed and predicted values partition groups.

The LQ had higher mean error (0.18 – 0.24) ± 0.17 compared with the MG (-0.03 – 0.22 ± 1.40. In the LQ mean error increased with more available data in the GT group whereas the reverse was observed for the MG. Residual mean square was similar in both models across sampled data. For LQ, RMS increased with more available data for the L60 group but increased for MG in the same direction of the G60 group. The Wald-Wolfowitz run test for the LQ was similar to those observed in the other models. None of the lactations showed a non-random error distribution for the MG model. The highest correlations between observed and predicted values were observed for both forms of the LQ and MG models.

Insert table 6 and Table 7 here

DISCUSSION

A desirable lactation model should be parsimonious, capable of depicting the production pattern and flexible enough to account for the influence of environmental factors affecting the curve shape without compromising accuracy. In addition, it should fit data from short lactations or standard test-day records (Pollott and Gootwine 2000). The LQ possesses these attributes. Since the performance of the LQ, LQ_n, BC and MG was not adversely affected by sampled data, either the L60 or G60 group of these models should be preferred to the IG in fitting shorter lactations which might be an advantage in pasture-based systems. These attributes of the LQ model and the better fitting throughout lactation with fewer parameters than the MG models suggests its comparative advantage and suitability for modeling lactation in dairy cows. Limitations of the new model are that it tends to emphasize the inflection points of the lactation curve in having parameters that determine both peak milk yield and day at peak. Secondly, like the other models, post-peak milk yield lactation especially around 200 DIM was not accurately predicted. It was not clear whether this feature is related to milk yield pattern in pasture-based cows or an artifact of the model properties. In autumn-calving systems, pasture-based Holstein-Friesian cows have been reported to exhibit a second post-peak milk yield in response to turning to pasture (Garcia and Holmes 2001).

Fitting Lactation Models to Average Lactation

Although daily milk yield in pasture-based dairy cows are subject to various perturbations such as seasonality of weather affecting pasture availability and quality, calving pattern and strategic nutrition management, these changes tend to even out at average herd level, hence the similarities in the goodness of fit of most of the models with the AL data. According to Pollott and Gootwine

(2000), any model that will improve on the IG model should be capable of representing early and peak milk yields more accurately than the IG model. Linear transformation of the test-day milk yield (LQ, MG and IG_L) yielded more randomly distributed residuals and improved the goodness of fit than previously suggested (Cobby and Le Du 1978, Pollott and Gootwine 2000) with respect to the IG and MG models respectively. It was noted that the four best-performing models in this study all had log-transformed components in their mathematical functions. The MG, PL, IG_L and LQ models achieved better prediction accuracy essentially through improvement in the prediction of early (except PL) and post-peak lactation milk yields, as opposed to the limitation of inaccurate peak milk yield prediction associated with empirical models (Olori et al. 1999, Pollott and Gootwine 2000, Macciotta et al. 2005, Sylvestre et al. 2006). Similar residual mean squares (RMS) values in all but the PE and QP models suggest little differences in the accuracy of prediction of the models when fitted to average lactation. However, lower Bayesian Information Criteria (BIC) in the log-transformed (LQ, MG and IG_L), and the mechanistic (BC, DJ and PT) models shows that these models achieved better prediction than the others. The flexibility of the PL, the Legendre polynomial (LEG) and cubic spline (SPL) functions enhanced better goodness of fit (Sylvestre et al. 2006).

Fitting Lactation Curves to Individual Cow's Lactation

Variability in individual cow's lactation data can influence the overall model performance (Olori et al. 1999). None of the models except the IG (EXLO = 3.10) predicted zero milk yield. Setting of lactation at test-day 1 to zero is a well reported (Tekerli et al. 2000, Silvestre et al. 2006) limitation of the IG model. The LQ ad MG achieved overall better performance than the models fitted directly to milk yield due to lower residuals occasioned by the log transformation.

The occurrence of *atypical* lactation pattern in individual cows, which may partly account for poor model fits, can be up to 30% (Olori et al. 1999, Macciotta et al. 2005, Silvestre et al. 2006). The occurrence of about 24% atypical lactations in our data might explain the poorer goodness of fit observed in fitting the models to individual cow's lactation. Similar results have been observed in previous studies (Silvestre et al. 2006, Berry et al. 2005). The higher residuals in the G60B and G60C, shows that the performance of the IG is influenced by availability of data in early lactation especially the first 60 days. The result corroborate the suggestion by Pollott and Gootwine (2000) that the poor goodness of fit of many models to individual cow's lactation was due to the poor fitting of the incline to peak yield part of the model either due to late first recorded test-day or paucity of data in early lactation i.e. L60A.

CONCLUSION

Lactation models were evaluated for goodness of fit to average and individual cow's lactation in pasture-based dairy system. The parameters of the model remain as previously explained. This model with biologically interpretable parameters, is easy to fit to either average or individual cow lactation data, and in contrast to the polynomial (PL, Ali and Schaeffer 1999) and the MG (Morant and Gnanasakthy 1989) models, it has three parameters that achieve a similar goodness of fit. The unique feature of the model that makes it suitable for diverse lactations modelling is the parameter *a* which fits either an inclining or declining lactation rate from initial milk yield, thus being less constrained by *a priori* assumption of an incline to peak yield. In addition, fitting the model parameters to the log transformed milk yield and days-in-milk minimizes the residuals and improves goodness of fit. The lactation models tested in this study adequately fitted dairy cow's average data although the PL, MG, IG_L and LQ produced more precise lactation parameters. The LQ and IG_L are parsimonious in utilising fewer parameters to achieve the same level of fitness as the PL and MG models. The BC model achieved an all round better goodness of fit than the other mechanistic models evaluated in this study. The best of the five selected models in fitting individual pasture-based Holstein-Friesian cow lactation was the LQ followed by the MG. It is necessary to routinely model lactation pattern of individual cows for management and breeding decisions. This study confirmed the influence of day at first test day and number of recorded test-days on the goodness of fit of lactation models. The new model was robust in fitting average or individual cow lactation and is recommended for fitting test-day milk yield in dairy systems. Further testing of the properties of the LQ model and its application to modelling data from other production systems are necessary to determine the robustness of the model.

ACKNOWLEDGEMENTS

This research was funded by a PhD research scholarship from DairyAustralia (DA), Tasmanian Institute of Agricultural Research (TIAR) and the University of Tasmania (UTAS). The authors gratefully acknowledge DA, TIAR and UTAS. Our gratitude also goes to Peter Nish and the management of TasHerd, Hadspen Tasmania, the staff of the dairy research farm in Victoria and the farmers who provided access to their data.

REFERENCES

Australian Dairy Herds Improvement Scheme (ADHIS) 2008. The Benefits of Herd Recording. Technical Report 7, ADHIS Pty Limited, September 2008, ver 1.

Ali, T. E., and L. R. Schaeffer. 1987. Accounting for covariances among test day milk yields in dairy cows. Can. J. Anim. Sci. 67**:**637-644.

Beever, D. E., A. J. Rook, J. France, M. S. Danoa and M. Gill. 1991. A review of empirical and mechanistic models of lactation performance by the dairy cow. Livest. Prod. Sci. 29:115-130.

Berry, D. P., V. E. Olori, A. R. Cromie, R. F. Veerkamp, M. Rath, and P. Dillon. 2005. Accuracy of predicting milk yield from alternative milk recording schemes**.** Anim. Sci. 80**:**53-60.

Cobby, J. M., and Y. L. P Le Du. 1978. On fitting curves to lactation data. Anim. Prod. 26:127-133.

Collard, B.L., P. J. Boettcher, J. C. M. Dekkers, D. Peticlerc, and L.R. Schaeffer. 2000. Relationships between energy balance and health traits of dairy cattle in early lactation. J. Dairy Sci. 83:2683-2690.

Congleton, J. R., and R. W. Everett. 1980. Application of the incomplete gamma function to predict cumulative milk production. J. Dairy Sci. 63:109–119.

Constantinides, A. 1988. Applied Numerical Methods with Personal Computers. 2nd ed. McGraw-Hill, New York, NY.

Dairy Australia 2008. Australia Dairy Industry in focus 2007: http://www.dairyaustralia.com.au/content/view/33/48/. Accessed on 31st March 2008.

Dave, B. K. 1971. First lactation curve of Indian water buffalo. *JNKVV* Res. J. 5: 93.

Dijkstra, J., J. France, M. S. Dhanoa, J. A. Maas, M. D. Hanigan, A. S. Rook, and D. E. Beever. 1997. A model to describe growth patterns of the mammary gland during pregnancy and lactation. J. Dairy Sci. 80:2340–2354.

Ferguson, J. D., and R. Boston. 1993. Lactation curves analysis: Comparison of gamma function, polynomial and exponential methods towards a mechanistic model of milk production. J. Dairy Sci. 76, (Supp. 1):268 (Abstract).

Garcia, S. C. and C. W. Holmes. 2001. Lactation curves of autumn- and spring-calved cows in pasture-based dairy systems. Livest. Prod. Sci. 68: 189-203.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson. 2002. ASReml User Guide Release 1.0. VSN International Ltd., Hemel Hempstead, UK.

Green, P. J., and B. W. Silverman. 1994. Nonparametric regression and generalized linear models. A roughness penalty approach. Chapman & Hall, London, UK.

Grossman, M., and W. Koops. 2003. Modelling extended lactation curves of dairy cattle: a biological basis for the multi-phasic approach. J. Dairy Sci. 86:988-998.

Jensen, J. 2001. Genetic evaluation of dairy cattle using test-day models. J. Dairy Sci. 84:2803-2812.

Kirkpatrick, M., W. G. Hill and R. Thompson. 1994. Estimating the covariance structure of traits during growth and ageing, illustrated with lactation in dairy cattle. Genet. Res. (Cambridge) 64:57-69.

Knight, C. H., M. Peaker, and C. J. Wilde. 1998. Local control of mammary development and function. Rev. Reprod. 3:104-112.

Knight, C. H. and C. J. Wilde. 1993. Mammary cell changes during pregnancy and lactation. Livest. Prod. Sci., 35:3-19.

Kolver, E. S., and S. B. Muller.1998. Performance and nutrient intake of high producing Holstein cows consuming pasture or a total mixed ration. J. Dairy Sci. 81:1403–1411

Leonard, T., and J. S. J. Hsu. 2001. *Bayesian methods*. Cambridge University Press, Cambridge, UK.

Macciotta, N. P. P., D. Vicario and A. Cappio-Borlino. 2005. Detection of different shapes of lactation curve for milk yield in dairy cattle by empirical mathematical models. J. Dairy Sci. 88:1178–1191.

Morant, S. V., and A. Gnanasakthy. 1989. A new approach to the mathematical formulation of lactation curves. Anim. Prod. 49:151-162.

Neal, H. D., and J. M. H. Thornley. 1983. The lactation curve in dairy cattle: A mathematical model of the mammary gland. J. Agric. (Cambridge) 101:389- 400.

Olori, V. E., S. Brotherstone, W. G. Hill, and B. J. McGuirk. 1999. Estimating variance components for test day milk records by restricted maximum likelihood with random regression animal model. Livest. Prod. Sci. 61:53–63.

Papajcsik, I.A., and J. Bodero, 1988. Modelling lactation curves of Friesian cows in a sub-tropical climate. Anim. Prod., 47: 201–205.

Pollott, G. E. 2000. A biological approach to lactation curve analysis for milk yield. J. Dairy Sci. 83:2448-2458.

Pollott, G. E., and E. Gootwine. 2000. Appropriate mathematical models for describing the complete lactation of dairy sheep. Anim. Sci. 71:197-207.

Ratkowsky, DA 1990. *Handbook of Nonlinear Regression Models.* New York: Marcel Dekker.

SAS 2002. *User's Guide Version 9.1*. SAS Institute Inc., Cary, NC, USA.

SAS 2004. *Genetic analysis of complex traits using SAS.* Saxton A. Ed. SAS Institute Inc., Cary, NC, USA.

Schaeffer, L. R. 2004. Application of random regression models in animal breeding. Livest. Prod. Sci. 86:35-45.

Sherchand, L., R. W. McNew, D. W. Kellog, and Z. B. Johnson. 1995. Selection of a mathematical model to generate lactation curves using daily milk yields of Holstein cows. J. Dairy Sci. 78:2507–2513.

Sikka, L. C. 1950. A study of lactation as affected by heredity and environment. J. Dairy Res. 17:231-252.

Silvestre, A. M., F. Petim-Batista, and J. Colaço. 2006. The accuracy of seven mathematical functions in modelling dairy cattle lactation curves based on test-day records from varying sample schemes. J. Dairy Sci. 89: 1813-1821.

Tekerli, M., Z. Akinci, I. Dogan, and A. Akcan. 2000. Factors affecting the shape of Lactation Curves of Holstein Cows from the Balikesir Province of Turkey. J Dairy Sci. 83: 1381–1386.

Tozer, P. R., and R. G. Huffaker. 1999. Mathematical equations to describe lactation curves for Holstein-Friesian cows in New South Wales. Aust. J. Agric. Res. 50(3):431-440.

Val-Arreola, D., 1 E. Kebreab, J. Dijkstra, and J. France. 2004. Study of the lactation curve in dairy cattle on farms in Central Mexico. J. Dairy Sci. 87:3789-3799.

White, I. M. S., R. Thompson, S. Brotherstone. 1999. Genetic and environmental smoothing of lactation curves with cubic splines. J. Dairy Sci. 82:632-638.

Wilmink, J. B. M. 1987. Adjustment of lactation yield for age at calving in relation to level of production. Livest. Prod. Sci. 16:321-334.

Wood, P. D. P. 1967. Algebraic model of the lactation curve in cattle. *Nature* 216:164-165.

Table 1. Summary of the data used in the study

	Number						
	Year	Herd	Cows	Lactation	Parity[1]	Mean	SD
DATA1	2005 - 2007	154	9,505	76,760	18, 098, 21,578, 37,084	12.7	5.34
DATA2	1998 - 2005	-	2,138	19,987	5,753, 4,503, 11,831	15.3	8.67

[1]Number of Herds, Cows and Lactations in parities 1, 2 and >= 3

Table 2. Data groups with number of cows, mean and standard deviation of milk yield used for testing the effect of days to first test-day and number of observations on the goodness of fit of five of the best fitting lactation models.

	Days to first Test-day	Number of Test days		
DATA1		5	6-7	>7
	< 60	L60A	L60B	L60C
		407, 12.9 (3.6)	500, 12.8 (4.0)	500, 12.6 (3.7)
	> 60	G60A(127)	G60B(299)	G60C(291)
		127, 11.6 (3.7)	300, 10.6 (3.3)	292, 12.0 (3.2)
DATA2	< 60	-	L60B	L60C
			188, 18.2 (5.9)	1350, 19.2 (6.3)
	> 60	-	-	G60C(291)
				588, 18.9 (6.2)

[1]Data groups are:
L60A = First test-day < 60 and 5 observations/cow; L60B = First test-day < 60 and 6 or 7 observations/cow; L60C = First test-day < 60 and> 7 observations/cow; G60A = First test-day > 60 and 5 observations/cow; G60B = First test-day > 60 and 6 or 7 observations/cow; and G60C = First test-day > 60 and > 7 observations/cow;

Table 3. The source, number of parameters and mathematical representation of the lactation models used in the study.

Model	Abbrv[a]	Lactation Model	Source	N*
Log-quadratic	LQ	$LogY_t = a\,(b - \mathrm{Log}t)^2 + c$	New	3
Modified gamma	MG	$LogY_t = a - bt' + ct'^2 + d/t$	Morant and Gnanasakthy (1989)	4
Incomplete gamma linear[b]	IG_L	$LogY_t = \mathrm{Log}\,(a) + b\,\mathrm{Log}t - ct$	Wood (1967)	3
Log-quadratic$_n$[c]	LQ_n	$Y_t = exp\,(\mathrm{a}\,(\mathrm{b} - \mathrm{Log}t)^2 + \mathrm{c})$	New	3
Modified$_n$ Gamma[c]	MG_n	$Y_t = \exp\,(a - bt' + ct'^2 + d/t)$	Morant and Gnanasakthy (1989)	4
Incomplete gamma	IG	$Y_t = at^b\,exp^{-ct}$	Wood (1967)	3
Exponential	EXP	$Y_t = a + b\,exp^{-kt} + ct$	Wilmink (1987)	3
Polynomial	PL	$Y_t = a + b\,(t_1) + c\,(t_1)^2 + d\,(\mathrm{Log}t_2) + e\,(\mathrm{Log}t_2)^2$	Ali and Schaeffer (1987)	5
Quadratic polynomial	QP	$Y_t = a + bt + ct^2$	Dave (1971)	3
Parabolic Exponential	PE	$Y_t = a\,exp\,(bt - ct^2)$	Sikka (1950)	3
Bi-compartmental	BC	$Y_t = a\,exp^{-bt} + c\,exp^{-dt}$	Ferguson and Boston (1993)	4
Dijkstra	DJ	$Y_t = a\,exp\,(b\,(1 - exp^{-ct}) / c - dt)$	Dijkstra et al. 1997	4
Pollott	PT	$Y_t = (a / 1 + ((1 - b) / b)\,exp\,(-cn)))\,(1 / 1 + ((1 - d) / d)\,exp\,(-en))$	Pollott (2000)	5
Modified Pollott	PT_2	$Y_t = (a / (1 + k\,exp\,(-1\,(n))\,(2 - exp\,(et)$	Pollott (2000)	2
Legendre polynomial	LP	$Y_t = \sum_{i=0}^{n} \alpha_i \phi_i(w)$	Kirkpatrick et al. 1994	4
Cubic splines	SPL	$Y_t = a_i + b_i\,(t - t_i) + c_i(t - t_i)^2 + d_i\,(t - t_i)^3$	Green and Silverman (1994)	3

[a] Abbrv = Model abbreviations, b = Log-transformed IG, c = untransformed models, N* = number of model parameters

Table 4. Predicted parameter estimates, residual means squares and Bayesian information criteria of lactation models fitted to average lactation data of Holstein Friesian dairy cows.

Abbrv[1]	Model and curve parameter estimates [2]	RMS	BIC
LQ	$LogY_t = -0.086\ (3.501 - \mathrm{Log}t)^2 + 2.590$	0.14	-16.09
MG	$LogY_t = 2.696 - 0.0019t' + 0.00000269\ t'^2 - 1.945 / t$	0.17	-16.10
IG_L	$LogY_t = \log(11.21) + 0.060 * logt - 0.002t$	0.14	-16.09
LQ_n	$Y_t = exp\ (-0.084\ (3.497 - \mathrm{Log}t)^2 + 2.589)$	0.04	-29.95
MG_n	$Y_t = exp\ (2.699 - 0.074t' + 0.0016t'^2 - 0.193/t)$	0.04	-28.95
IG	$Y_t = 11.13t^{0.0626} exp^{-0.0021t}$	0.05	-26.02
EXP	$Y_t = 13.825 - 904.2\ exp^{-0.46t} + 0.018t$	0.03	-30.45
PL	$Y_t = -10.22 + 28.84t_1 - 10.29t_1^2 + 15.44 \ln t_2 - 2.76(\ln t_2)^2$	0.03	-36.72
QP	$Y_t = 13.264 - 0.012t - 0.00001t^2$	0.10	-19.82
PE	$Y_t = 4.33\mathrm{E}^{-90}\ exp\ (-0.130t - 0.100t^2)$	22.85	34.63
BC	$Y_t = 8.38\ exp^{-0.0014t} + 5.25 exp^{-0.0014t}$	0.17	-16.22
DJ	$Y_t = 0.502\ exp\ (11.938\ (1 - exp^{3.617t}) / 3.617 - 0.0014t)$	0.17	-16.22
PT	$Y_t = (1.0e^{-8} / 1 + ((1 - 0.083) / 0.083)\ exp\ (0.001n))$ $(1/1 + ((1 - 2.39e^{-62}) / 2.39e^{-62})\ exp\ (2.24e^{-62}n))$	0.20	-16.22
PT_2	$Y_t = (13.368 / (1 + 0.0001)\ exp\ (-1\ (t - 150)))\ (2 - exp(0.001031t))$	0.09	-19.53
[3]LEG	$Y_t = \sum_{i=0}^{n} \alpha_i \phi_i (w)$ [Tday = 0.055, Leg (Tday) = 4.059, 7.811, and 0.112]	0.06	-27.12
[3]SPL	$Y_t = a_i + b_i(t - t_i) + c_i(t - t_i)^2 + d_i(t - t_i)^3$ [Tday = -0.016, *mu* = 13.45, spl(Tday3) = -0.022]	0.07	-22.83

[1]Model Abbreviations are: LQ = Log-quadratic (new model), MG = Modified *gamma* (Morant and Gnanasakthy 1989), IG_L = modified Incomplete *gamma* (Wood 1967), LQ_n = (modified LQ), MG_n = modified MG, IG = Incomplete *gamma*, EXP = Exponential (Wilmink 1987), PL = Polynomial (Ali and Schaeffer 1987), QP = Quadratic Polynomial (Dave 1971), PE = Parabolic Exponential (Sikka 1950), BC = Bi-compartmental (Ferguson and Boston 1993), DJ = Dijkstra (Dijkstra et al. 1997), PT = Pollott (Pollott 2000), PT_2 = modified Pollott, LEG = Legendry polynomial (Kirkpatrick et al. 1994) and SPL = Cubic splines (Green and Silverman 1994).

[3] Parameters of the semi-parametric models in angled brackets

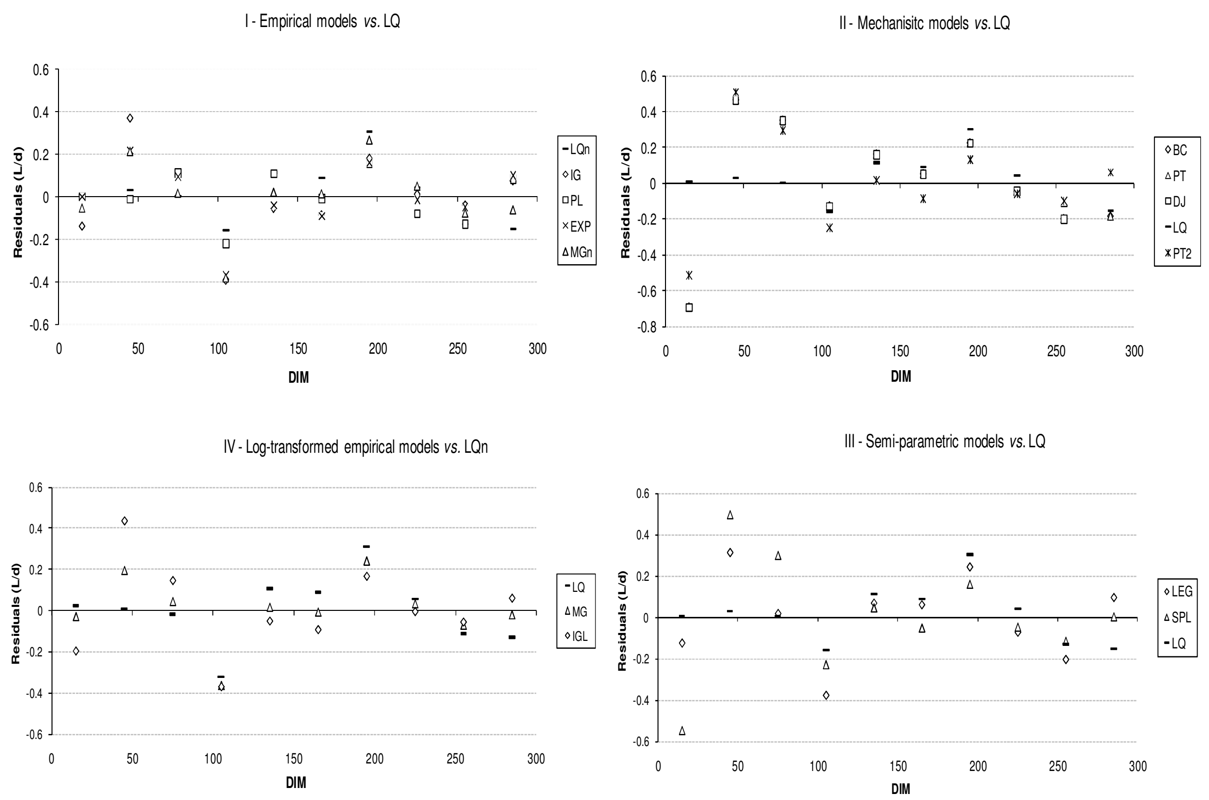

Figure 1. Comparison of plots of residuals against day in milk from fitting Average lactation (DATA1) to the Log-quadratic and other lactation models

I — LQn = modified log-quadratic; ◊ IG = Incomplete *gamma*; (Wood 1967); □ PL = Polynomial (Ali and Schaeffer 1987); x EXP = Exponential (Wilmink 1987) and Δ MGn = modified MG (Morant and Gnanasakthy 1989),

II ◊ BC = Bi-compartmental (Ferguson and Boston 1993); Δ PT = Pollott (2000); □ DJ = (Dijkstra et al. 1997); — LQ = log-quadratic and PT2 = modified Pollott.

III ◊ LEG = Legendre polynomial (Kirkpatrick et al. 1994); Δ SPL = cubic splines (Green and Silverman 1994) and — LQ = log-quadratic (untransformed).

IV — LQ = log-quadratic; Δ MG = modified *gamma* (Morant and Gnanasakthy 1989); and ◊ IG_L = Log-transformed Incomplete *gamma*, (Wood 1967).

Table 5. Predicted initial, peak, and 305d milk yields, day at peak and persistency of lactation obtained from fitting various lactation models to the Herd lactation data of pasture-based Holstein-Friesian cows.

Model Abbrev[1]	Predicted milk yield parameters[2]						
	Initial	Peak	Peak day	Nadir	305d Yield	Stdev	Persistency (mL/d)
Actual	9.73	14.2	21	7.98	3313	4.00	19.6
LQ	9.05	13.3	33	8.67	3312	3.95	18.1
MG	9.07	13.2	34	9.07	3309	3.96	18.3
IG_L	12.1	12.9	30	8.57	3334	3.95	16.9
LQn	9.16	13.3	33	8.79	3316	3.97	17.8
MG_n	8.29	13.2	32	10.6	3315	3.96	17.7
IG	12.0	12.9	31	8.65	3354	3.95	16.8
EXP	-130	13.4	22	-130	2945	31.8	18.0
PL	5.23	13.5	30	5.23	3303	4.21	18.0
QP	13.2	13.2	4	8.67	3351	4.04	15.3
PE	13.2	13.2	4	8.37	3322	4.22	16.6
BC	13.9	13.6	4	8.89	3341	4.13	15.2
DJ	13.6	13.6	4	8.89	3339	4.13	15.3
PT	12.8	12.8	4	9.46	3335	3.91	11.0
PT_2	13.4	13.4	4	12.9	3972	13.67	14.0
LEG	12.5	13.1	21	9.07	3326	4.00	16.5
SPL	13.4	13.4	4	8.54	3321	4.08	16.2

[1]Model Abbreviations are LQ = Log-quadratic (new model), MG = Modified *gamma* (Morant and Gnanasakthy 1989), IG_L = Log-transformed Incomplete *gamma*, (Wood 1967), LQ_n, MG_n, (untransformed), IG, EXP = Exponential (Wilmink 1987), PL = Polynomial (Ali and Schaeffer 1987), QP = Quadratic polynomial (Dave 1971), PE = Parabolic exponential (Sikka 1950) BC = Bi-compartmental (Ferguson and Boston 1993), DJ = Dijkstra et al. 1997), PT = Pollott (2000), MPT_1 = (4-parameter modified Pollott), MPT_2 = (2-parameter modified Pollott), LEG = Legendre polynomial (Kirkpatrick et al. 1994) and SPL = cubic splines (Green and Silverman 1994).

[2]Except for days at peak and nadir, all milk yield parameters are in litres.

Table 6. Comparison of the goodness of fit criteria of the Wood (IG), Bicompartmental (BC), Log quadratic polynomial (LQ_n and LQ), and Morant (MG) models fitted to individual cow's lactations in entire DATA1 and DATA2[1].

	Lactations	Test-days	Error	SD_e	RMS	SD_r	W	EXLO	EXHI	r
DATA1										
[2]IG	9, 502	76,761	0.001	0.024	16.9	33.92	121	3.10	0.03	0.18
BC	9, 502	76,761	-0.08	0.024	14.0	13.12	117	0.00	0.03	0.49
LQ_n	9, 502	76,761	0.001	0.024	10.1	8.73	147	0.00	0.03	0.51
LQ	9, 502	76,761	0.24	2.58	0.07	0.05	104	0.00	0.02	0.99
MG	9, 502	76,761	-0.01	‾1.50	0.07	0.07	0	0.00	0.03	0.99
DATA2										
IG	2,138	19,987	0.007	0.27	8.6	2.87	55	0.00	0.08	0.77
BC	2,138	19,987	-0.02	0.03	12.4	15.5	55	0.00	0.11	0.73
LQ_n	2,138	19,987	-0.007	0.03	8.80	10.51	55	0.00	0.08	0.77
LQ	2,138	19,987	0.15	0.20	0.03	0.05	57	0.00	0.09	0.99
MG	2,138	19,987	0.12	0.17	0.03	0.05	52	0.00	0.11	0.99

[1]DATA1 = Test-day data Tasmania; DATA2=Test-day data Victoria; Error = Mean residuals (difference between daily actual and predicted milk yield) per test-day for every cow; SDe = standard error of residuals ; RMS = Residual Mean Square (averaged for individual cow lactation; SDr = standard error of RMS; W = Wald-Wolfowitz test (number of lactations with random distribution); EXLO = percentage of estimated milk yields lower than or equal to zero; EXHI = percentage of estimated milk yields higher than 40 kg; r = correlation between actual daily and predicted milk yield**.**

[2]Lactation models are:
IG = Incomplete gamma (Wood 1967), $Y_{(t)} = a\, t^{b} e^{-ct}$
BC = Bi-compartmental model (Ferguson and Boston 1993), $Y_t = a\, e^{-bt} + de^{-ct}$
LQ_n = Log-quadratic (New model), $Y_t = exp\,(a\,(b - \text{Log}\, t)^2) + c$
LQ = Log-quadratic (New model), $Log\ Y_{(t)} = a\,(b - \text{Log}t)^2 + c$
MG = Modified gamma (Morant and Gnanasakthy (1989), $LogY_t = a - bt' + ct'^2 + d/t$

Table 7. Comparison of the goodness of fit criteria of the Wood (IG), Bicompartmental (BC), Log quadratic polynomial (LQ_n and LQ), models fitted to data samples based on differences in days to first test-day and number of observations per cow in DATA1

	Lactations	Test-days	Error	SD_e	RMS	SD_r	Wolf	EXLO	EXHI	R^2
IG										
[2]L60A	408	2035	0.43	2.01	14.9	23.50	54	4.62	0.00	0.89
L60B	500	3308	0.29	1.55	15.0	29.85	38	3.23	0.09	0.11
L60C	500	4424	0.26	1.55	13.6	27.92	5	2.82	0.04	0.28
G60A	128	635	0.47	2.05	20.7	55.33	23	5.51	0.00	0.29
G60B	300	1961	0.57	2.19	17.1	41.35	1	6.78	0.00	0.24
G60D	292	2469	0.57	2.33	18.8	45.2	0	5.14	0.00	0.12
BC										
L60A	408	2035	-0.005	0.125	25.9	26.89	53	0.00	0.00	0.55
L60B	500	3308	-0.006	0.015	13.11	12.08	30	0.00	0.09	0.53
L60C	500	4424	-0.008	0.017	13.3	10.23	3	0.00	0.04	0.46
G60A	128	635	-0.004	0.014	16.7	13.74	25	0.00	0.00	0.58
G60B	300	1961	$-9.0e^{-4}$	0.018	10.5	11.98	5	0.00	0.00	0.52
G60D	292	2469	-0.005	0.016	11.9	10.27	0	0.00	0.00	0.48
LQn										
L60A	408	2035	0.002	0.019	15.3	15.02	69	0.00	0.00	0.47
L60B	500	3308	0.002	0.021	9.9	8.62	48	0.00	0.09	0.51
L60C	500	4424	$8.1e^{-5}$	0.021	10.1	7.83	8	0.00	0.02	0.53
G60A	128	635	0.003	0.016	10.6	8.95	17	0.00	0.00	0.50
G60B	300	1961	0.003	0.122	8.1	10.52	4	0.00	0.00	0.49
G60D	292	2469	-0.004	0.024	93.03	8.19	0	0.00	0.00	0.50
LQ										
[2]L60A	408	2035	0.23	0.20	0.09	0.09	53	0.00	0.00	0.99
L60B	500	3308	0.19	0.13	0.07	0.05	10	0.00	0.15	0.99
L60C	500	4424	0.26	0.17	0.07	0.06	0	0.00	0.05	0.99
G60A	128	635	0.18	0.14	0.08	0.06	14	0.00	0.00	0.99
G60B	300	1961	0.20	0.14	0.09	0.05	3	0.00	0.00	0.99
G60D	292	2469	0.24	0.21	0.07	0.07	0	0.00	0.00	0.99
MG										
L60A	408	2035	0.04	1.49	0.14	0.15	0	0.00	0.00	0.99
L60B	500	3308	-0.03	1.33	0.03	0.06	0	0.00	0.15	0.99
L60C	500	4424	0.12	1.54	0.07	0.05	0	0.00	0.05	0.99
G60A	128	635	0.22	1.31	0.12	0.11	0	0.00	0.00	0.99
G60B	300	1961	-0.05	1.04	0.08	0.06	0	0.00	0.00	0.99
G60D	292	2469	0.03	1.39	0.07	0.07	0	0.00	0.00	0.99

[1]DATA1 = Test-day data Tasmania; DATA2=Test-day data Victoria; Error = Mean residuals (difference between daily actual and predicted milk yield) per test-day for every cow; SD_e = standard error of residuals; RMS = Residual Mean Square (averaged for individual cow lactation; SD_r = standard error of RMS; W = Wald-Wolfowitz test (number of lactations with random distribution); EXLO = percentage of estimated milk yields lower or equal to zero; EXHI = percentage of estimated milk yields higher than 40 L, R^2 = correlation between actual daily and predicted milk yield.

[2]Data groups are:
L60A = First test-day < 60 and 5 observations/cow; L60B = First test-day < 60 and 6 or 7 observations/cow; L60C = First test-day < 60 and> 7 observations/cow; G60A = First test-day > 60 and 5 observations/cow; G60B = First test-day > 60 and 6 or 7 observations/cow; and G60C = First test-day > 60 and > 7 observations/cow.

[2]Lactation models are:
IG = Incomplete gamma (Wood 1967), $Y_{(t)} = a\, t^{b}\, e^{-ct}$
BC = Bi-compartmental model (Ferguson and Boston 1993), $Y_t = a\, e^{-bt} + de^{-ct}$
LQ_n = Log-quadratic (New model), $Y_t = exp\,(a\,(b - \mathrm{Log}\ t)^2) + c$
LQ = Log-quadratic (New model), $Log\ Y_{(t)} = a\,(b - \mathrm{Log}t)^2 + c$
MG = Modified gamma (Morant and Gnanasakthy (1989), $LogY_t = a - bt' + ct'^2 + d/t.$